A Comparison of Four Schemes for Codification of Problem Lists

James R. Campbell M.D., University of Nebraska Medical Center, Omaha NE
Thomas H. Payne M.D., Group Health Cooperative of Puget Sound, Seattle, WA

ABSTRACT

We set out to evaluate the completeness of four major coding schemes in representation of the patient problem list: the Unified Medical Language System (UMLS, 4th edition), the Systematized Nomenclature of Medicine (SNOMED International), the Read coding system (version 2), and the International Classification of Diseases (9th Clinical Modification)(ICD-9-CM). We gathered 400 problems from patient records at primary care sites in Omaha and Seattle. Matching these against the best description found in each of the coding schemes, we asked five medical faculty reviewers to rate the matches on a five-point Likert scale assessing their satisfaction with the results. For the four schemes, we computed the following rates of dissatisfaction, satisfaction, and average scores:

	Unsatisfactory	*Satisfactory*	*Average*
UMLS	*.22*	*.65*	*3.92*
SNOMED	*.24*	*.60*	*3.57*
READ	*.38*	*.38*	*2.99*
ICD-9-CM	*.42*	*.37*	*2.85*

From this analysis, we conclude that UMLS and SNOMED performed substantially better in capturing the clinical content of the problem lists than READ or ICD-9-CM. No scheme could be considered comprehensive. Depending on the goal of systems developers, UMLS and SNOMED may offer different, and complementary, advantages.

INTRODUCTION

The 1991 publication by the Institute of Medicine (IOM)[1] has set the goals and defined the agenda for this decade of computerized patient record (CPR) development. Critical issues for research into CPR design include data standards for content and clinical vocabulary. Criteria for selection of the best schemes or strategies are not a matter of general agreement nor have comprehensive studies been done to define those attributes which best guarantee utility of a proposed data standard.

A definition of the CPR content and vocabulary must begin, practically speaking, with studies of the data recorded in current paper record systems. Virtually all published studies[2-5] have focused on limited clinical realms, and on one or two coding schemes. Recent work accomplished by the Codes and Structures work group of the Computerized Patient Record Institute (CPRI)[6] has added meaningful information to the discussion of a core data scheme. This study evaluated eight major coding systems, comparing them against the content of clinical documents garnered from four medical centers. The best schemes emerging from this evaluation were the Systematized Nomenclature of Medicine (SNOMED International)[7,8] and the Unified Medical Language System (UMLS, 4th edition)[9,10] of the National Library of Medicine. One limitation of this study was the exclusion of tabular or flow-charted source material, much of which is critical to the content of the patient record.

Probably the best studied and most useful feature of medical records is the problem list [11,12]. Established as a required feature of hospital records by the Joint Committee for the Accreditation of Hospitals (JCAHO), it has been shown to improve coordination of care, assure follow-up of patient concerns, and better organize record keeping. Nonetheless, problem lists were excluded from the CPRI study which only evaluated textual data.

We were concerned that such a shortcoming should be corrected, and decided to evaluate the problem coding capability of the best of the coding schemes studied by the CPRI. We limited our project to those schemes that were best oriented toward diagnostic findings. These included SNOMED, UMLS, the READ coding scheme (version 2)[13] employed by the National Health Service of the United Kingdom, and the International Classification of Diseases (9th Clinical Modification)(ICD-9-CM). As a part of planning projects in our respective institutions, this served the further purpose of refining our plans for local CPR implementation.

0195-4210/94/$5.00

METHODS

In order to prepare a study that would be geographically representative, we obtained medical records material from the University of Nebraska Medical Center (UNMC) Internal Medicine clinics, and from the Primary Care clinics of the Group Health Cooperative of Puget Sound (GHC). A list of candidate problems taken from medical records at each institution was assembled. The 200 most frequent problems found in COSTAR®[14] records at UNMC were identified by record query. Charts were pulled by convenience sample from pediatric records at GHC clinics. The first 200 problems found in these records were added to the list. From the total of 400 candidates, duplicate items were eliminated and ambiguous terms were clarified by chart review. This yielded a study sample of 359 problems taken from adult and pediatric medicine at the two institutions.

We matched the study sample against the four major coding schemes using textual references and coding browsers. In some cases, no reasonable match for the concept could be found in the coding scheme using the most liberal clinical judgement. These items were scored zero. In some cases exact lexical matches were found. These were scored as five points. All items for each coding scheme with "near matches" were assembled on a study sheet, comparing the source problem and the nearest coding match. The sheets were prepared using the published terminology of the coding scheme, with no attempt at ordering or clarification of terminology. For example, the problem "back pain" was displayed opposite the four coding schemes as in figure 1.

Five university faculty from the two institutions, all users of the problem oriented record, were asked to rate each match using the five point Likert scale listed in figure 1. This was based upon previous work at GHC and was designed to reflect their satisfaction with the coded representation of the original problem.

The results of the five evaluations were collated and frequency statistics were computed for match scores. Average acceptability scores were computed for each coding scheme. A review of this summary made it clear that modifying words used in the publication of each scheme had an impact on the faculty evaluation. For example, phrases including words such as "unspecified" and "NOS" were scored lower by faculty physicians, even when the conceptual content was otherwise exact. We did a second lexical analysis by stripping these words from the codes. For example, we converted "BACK, NOS; PAIN, NOS" to "BACK PAIN." We then recomputed the frequency of exact matches for each scheme using this revised code list.

```
BACK PAIN  1 2 3 4 5     BACK, NOS; PAIN, NOS     (SNOMED)
                         BACKACHE                 (READ)
                         BACK PAIN                (UMLS)
                         BACK ACHE UNSPECIFIED    (ICD-9-CM)

1=EXTREMELY DISSATISFIED
2=QUITE DISSATISFIED
3=NEUTRAL, NOT SATISFIED BUT NO OBJECTION
4=QUITE SATISFIED
5=EXTREMELY SATISFIED
```

Figure 1

RESULTS

From the original list of 400 problems, 359 remained after duplications and ambiguous terms were discarded. Misprints on coding sheets caused one or two terms to be ignored in the final analysis of each scheme. Table 1 summarizes the frequency of scores assigned to each scheme by the five faculty evaluators. The column at far right is the average acceptability score for all problems and all evaluators.

A review of the scores assigned to each concept made it clear that our presentation biased the faculty against schemes which used coding parlance such as "NOS." On the other hand, a programmer might choose to implement those schemes having stripped such repetitive terms from the vocabulary presented to users of the computerized record. Table 2 summarizes the exact matches for each scheme as they were taken from their published source (baseline). The far right column of table 2 summarizes the exact match score after stripping terms such as "NOS" and "unspecified" and reanalyzing for exact matches ignoring punctuation and lexical order.

DISCUSSION

For a coding scheme to be useful in development of a CPR, it must be comprehensive, multi-disciplinary, concise, provide meaningful taxonomic relationships, be linguistically representative, and should support useful links to administrative and knowledge schemes. The purpose of this project was to evaluate the completeness of four candidate coding schemes relative to the conceptual content of the problem list at our institutions. We chose schemes that performed well in other evaluations (SNOMED, UMLS, READ), or are in common use in the United States (ICD-9-CM).

The accuracy of a coded representation of a clinical concept may be judged in a variety of ways, but we chose to echo the coded output to senior clinicians and asked them to compare the results to those originally recorded. The lexical method we used to prepare the evaluation may have created a bias against schemes which are conceptually and taxonomically oriented. This same method may have favored those that are linguistically based. The frequency of exact

Table 1
Scoring Frequency and Average Scores by Scheme

	0	1	2	3	4	5	Average
UMLS	.00	.08	.14	.14	.07	.58	3.92
SNOMED	.02	.15	.07	.16	.20	.40	3.57
READ	.03	.14	.21	.23	.24	.15	2.99
ICD-9-CM	.01	.21	.20	.22	.11	.26	2.85

Table 2 Matches at Baseline and After Stripping of Modifiers

	Baseline	Modifiers stripped
	Exact match	Exact match
UMLS	.52	.52
SNOMED	.10	.38
READ	.13	.21
ICD-9-CM	.09	.10

matching with UMLS relative to all other schemes may be an example of this bias. This should be kept in mind when evaluating the results.

Reviewing the summary statistics makes it clear that UMLS is a well developed and reasonably complete scheme with some obvious advantages for problem list encoding. If a system developer is interested in populating a data dictionary, and establishing links to systems such as ICD-9-CM, MESH and SNOMED, then UMLS will provide many benefits by virtue of its cross-references. If the developer is further interested in a gateway from the problem list to the medical literature, then UMLS is the only choice. That is the purpose for which it was designed.

On the other hand, the internal structuring of UMLS is primarily semantic, which may support natural language processing but is not helpful for deducing relationships between concepts that may be important for decision support. SNOMED is designed to support such a taxonomy, and is also cross referenced into ICD-9-CM, International Classification of Diseases for Oncology (ICD-O), and American Hospital Formulary (AHFS) coding schemes. The multi-axial features of the scheme are especially helpful to avoid an explosion of the number of coding elements when faced with adding new terms with common features. Based upon work done by the Major Codes group of the CPRI, SNOMED is also much more complete outside of the domain of diagnoses, making it a more suitable candidate if the system developer is interested in a coded problem list with meaningful relationships to other coded portions of the CPR. Nonetheless, a systems developer who chooses SNOMED must be prepared to build a clinician term vocabulary, an effort in itself that requires substantial work.

Comparing the cumulative frequencies of unsatisfactory matches by coding scheme (scores 0-2), it is clear that the READ and ICD-9-CM schemes perform much more poorly for problem coding. They showed 38% and 42% unsatisfactory matches respectively, approximately twice the rate for either UMLS or SNOMED. This performance generally mirrors a weaker showing by these schemes in other CPR domains[6].

In summary, the systems developer of a CPR has many choices to make when choosing standard coding schemes. Based upon this evaluation of problem encoding, both UMLS and SNOMED are more complete than alternative systems. Depending upon the goals of the project, UMLS will offer advantages for systems that wish to employ natural language processing and literature links. If the goal is an integrated, comprehensive coded record with decision support features, SNOMED is more attractive.

Reference

[1]. Dick RS, Steen EB, eds. The Computer-Based Patient Record: An Essential Technology for Health Care. 1991; Institute of Medicine, National Academy of Sciences

[2]. Henry SB, Campbell KE, and Holzemer WL. Representation of Nursing Terms for the Description of Patient Problems Using SNOMED III. Annu Symp Comput Appl Med Care 1993; In press

[3]. Zielstorff RD, Cimino C, Barnett GO, et al. Representation of Nursing Terminology in the UML Metathesaurus: A Pilot Study. Proc Annu Symp Comput Appl Med Care. 1992; 354-348

[4]. Campbell JR, Kallenberg GA, Sherrick RC. The Clinical Utility of META: An Analysis for Hypertension. Proc Annu Symp Comp Appl Med Care 1992; 397-401

[5]. Payne TH, Martin DR. How Useful is the UMLS Metathesaurus in Developing a Controlled Vocabulary for an Automated Problem List? Proc Annu Symp Comput Appl Med Care. 1993; 705-709

[6]. Chute CG, Cohn S, Campbell JR, Campbell KE, Oliver DE and the CPRI Major Codes Workgroup. The Adequacy of Existing Clinical Classifications: Content Coverage. Submitted for publication

[7]. Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, eds. The Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. 1993, College of American Pathologists

[8]. Cad RA; Robboy S. Progress in Medical Information Management. Systematized Nomenclature of Medicine (SNOMED). JAMA 1980; 243(8):756-762

[9]. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med 1993; 32:281-291

[10]. Lindberg DAB, Humphreys BL, McCray AT. Yearbook of Medical Informatics. "The Unified Medical Language System" 1993; 41-51

[11]. Weed,L. Medical Records that Guide and Teach. NEJM 1968; 278:593-600

[12]. Weed L. The Problem Oriented record as a Basic Tool in Medical Education, Patient Care and Research. Ann Clin Res 1971; 3(3):131-4

[13]. Read Codes File Structure Version 3: Overview and Technical Description. Woodgate, Leicestershire, UK: NHS Centre for Coding and Classification, 1993

[14]. Computer Stored Ambulatory Record. Research Digest Report of NCHSR Grant HS-00240. Barnett GO. United States DHHS Publication HRA 76-3145, National Center for Health Services Research, 1976